Air Quality Final Project

Qianyue Wang, Zijing Zhang, Ethan Louie, Nate Hiatt

01

Business Problem

- A need for real-time air quality monitoring and predictive estimation of pollutants in urban environments.
- This model aims to solve the problem of estimating Carbon Monoxide (CO) concentration based on other readily available sensor readings like Nitrogen Dioxide (NO2)

02

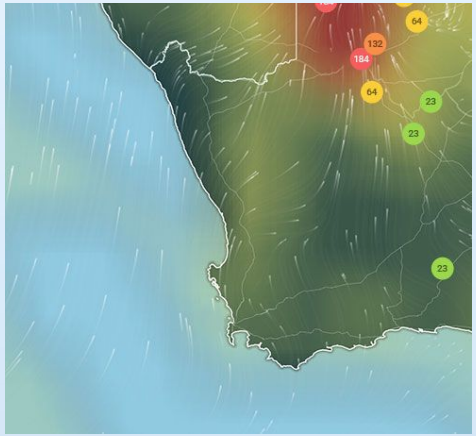
Architecture

Google Cloud

- **Pipeline** → IPYNB → GCS + Pub/Sub → Dataflow → BQ/BQML→ Looker
- **Security** → granting the Pub/Sub Service Account minimal permission required to write data. All external keys should be moved to Secret Manager for Prod.
- **Failure handling** → Pub/Sub DLT to isolate stream failure from malformed data records.
- **Operational and Cost Efficiency** → **Serverless** (Cloud Functions, Pub/Sub), BigQuery Time Partitioning on the historical data → reduces query costs.

03

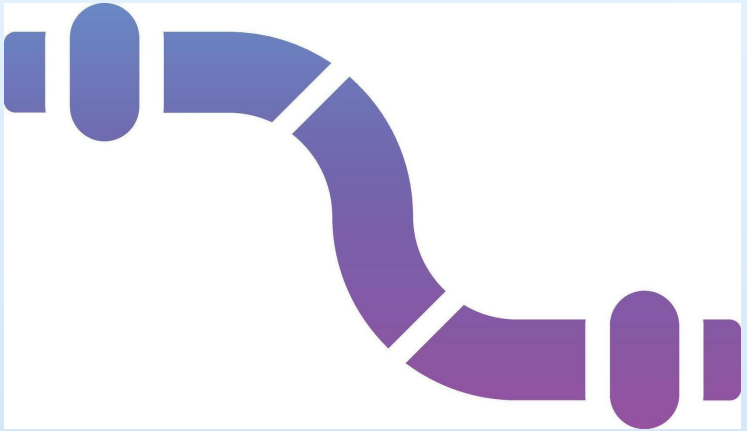
Batch Pipeline



- Objective: Establish a robust Data Warehouse using historical sensor data.
- Air Quality Dataset: Sourced (2004–2005) from Kaggle. Linked in the **GitHub**

04

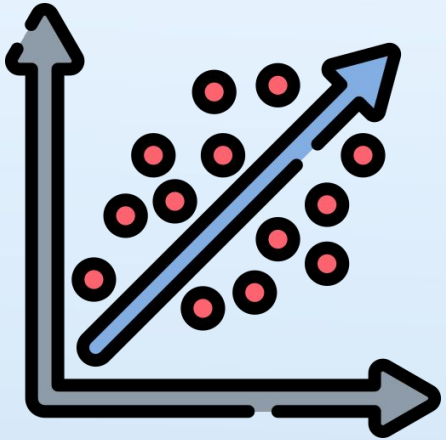
Stream Pipeline



- Objective: Build a serverless pipeline to capture live environmental data.
  - API Integration: Interfaced with the Open-Meteo API to fetch live Carbon Monoxide, NO2, and Ozone readings for Rome, Italy.
- **GitHub**

05

Regression Model



- **Label**: CO_2 Levels
- **Features**: Tungsten/Benzene/Nitrogen Levels, City, Population

```
--- Model Metrics ---
R2 Score: 0.7393
Mean Absolute Error: 0.5350 mg/m^3
Mean Squared Error: 0.5414
```

06

Looker Dashboard

- Dashboard Link: Here
- Key Insights: Skewed Tungsten Distribution, City has minimal effect on the model

07

Business Impact

- Significant for environmental agencies, public health organizations, and smart city initiatives. Provides near real-time insights into pollutant levels.
- Faster decisions on issuing air quality warnings, managing traffic flow, or deploying intervention strategies, ultimately leading to improved public health outcomes and regulatory compliance.

Thank You